Use of Controlled Vocabularies

Potential applications to time series data

It's 1984 ...

http://4.bp.blogspot.com/-gQWSBV6EnBo/TiRcN9unwbI/AAAAAAAAPXs/tydq754TRoI/s1600/vgdvd3.jpg

Friday night

http://i.telegraph.co.uk/multimedia/archive/01594/corey-haim_1594267c.jpg

MISSION: **PIZZA**

"What's their address?"

"1428 Elm St."

"Be there in 30 mins or less
...or it's free."

http://img2.timeinc.net/people/i/2006/celebdatabase/patrickdempsey/patrick_dempsey5_180_240.jpg
http://img.timeinc.net/time/daily/2009/0901/360_elm_street_0105.jpg

MISSION: **MOVIE**

“GET THE ONE WITH…”

- Righteous **girl** + bodacious **dude**
- **Light beams** flying everywhere
- The one with the “**Force**”

http://swhomevid.site40.net/vhs_anh1.jpg

The search begins...

https://deadlymovies.files.wordpress.com/2012/01/photo-2.jpg

MOVIE: **FOUND!**

- ✓ Righteous **girl** + bodacious **dude**
- ✓ **light beams** flying everywhere
- ✓ The one with the "**Force**"

http://www.chud.com/wp-content/uploads/2012/07/MEGAFORCE-FRONT-copy.jpg

BACK AT THE HOUSE...

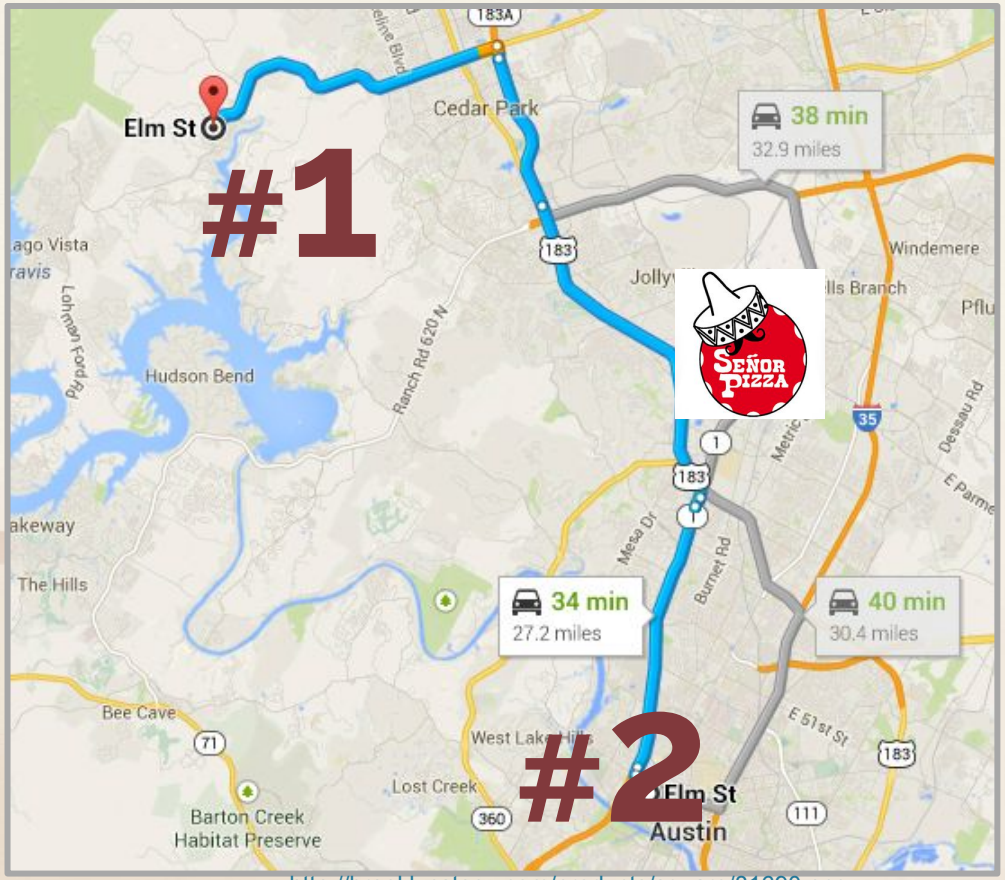
http://static.guim.co.uk/sys-images/Film/Pix/pictures/2010/2/24/1267010697702/Winona-Ryder-Kim-Walker-L-001.jpg

A Tale of Two Elm Streets

Which ‘Elm St’ was that?

http://i.huffpost.com/gen/1594554/thumbs/o-PIZZA-facebook.jpg
http://themaindamie.files.wordpress.com/2013/10/a-nightmare-on-elm-street-2010-horror-movies-11384409-720-297.jpg
http://img.timeinc.net/time/daily/2009/0901/360_elm_street_0105.jpg

ELM ST. near Austin, TX



http://brooklynsteez.com/products/square/81690.png
http://tinyurl.com/mtmlbpx

At least, you have MEGAFORCE

- ✓ Righteous **girl** + bodacious **dude**
- ✓ **light beams** flying everywhere
- ✓ The one with the "**Force**"

http://www.chud.com/wp-content/uploads/2012/07/MEGAFORCE-FRONT-copy.jpg

Betamax vs. VHS

http://images.sodahead.com/slideshows/000000228/betamax-11979413784_xlarge.jpeg
http://upload.wikimedia.org/wikipedia/commons/thumb/3/3c/Betavhs2.jpg/220px-Betavhs2.jpg

WHAT HAPPENED, DUDE?

PIZZA
no results = ***ambiguous*** address

MEGA FORCE
got results, not what was ***expected***

BETAMAX
results weren't ***useful***

F.A.I.R.

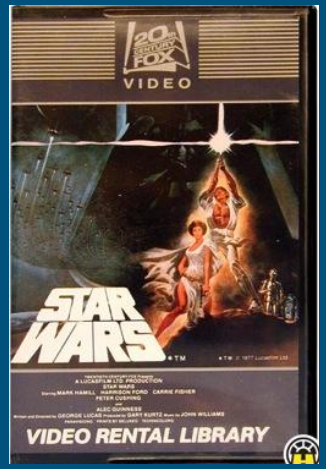
Findability and Accessibility has greatly improved.

Scaling interoperation is difficult

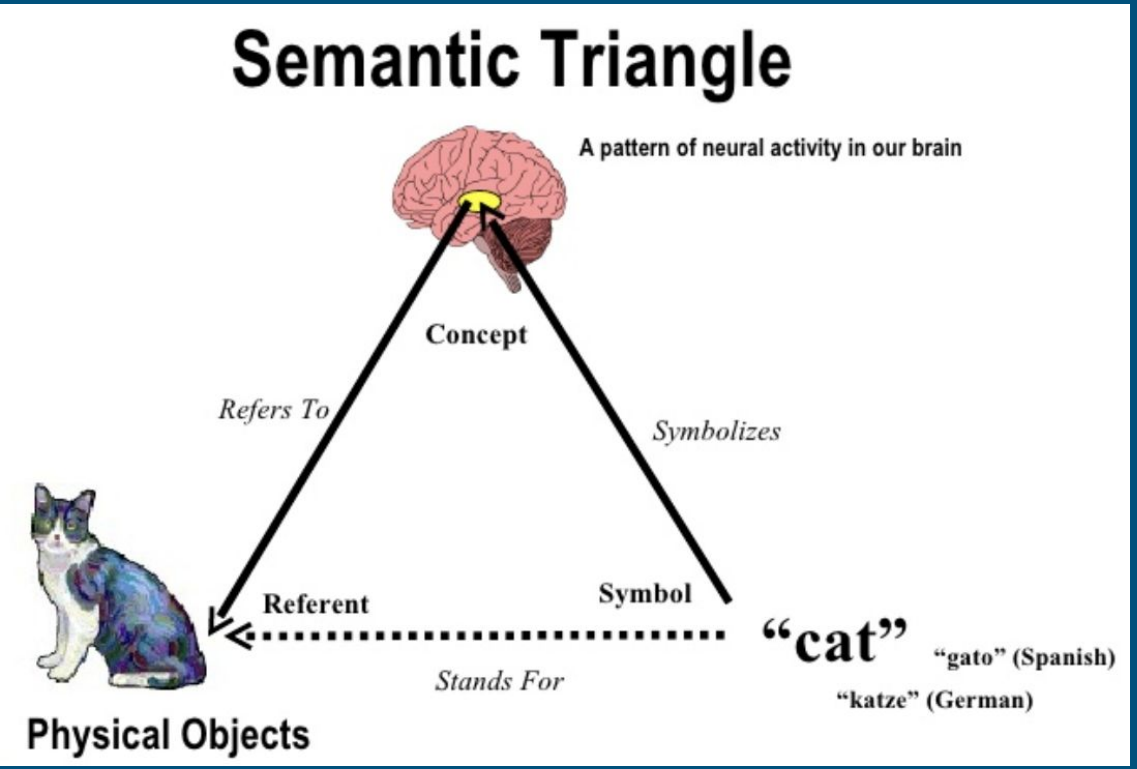
Without progress, ocean data risks being siloed

What does interoperation require?

AGREEMENT



Of Concepts & Computers



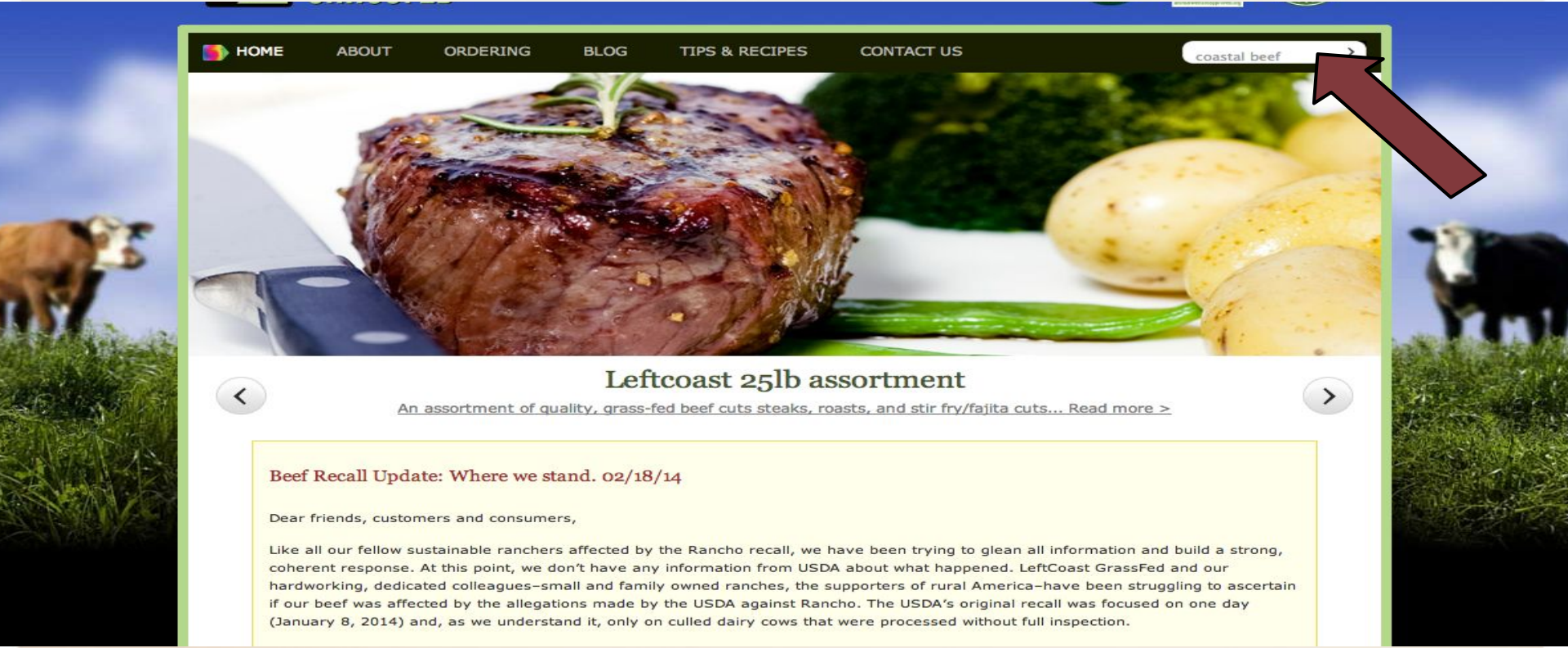
Referents MUST pass through Concepts

Computers can only exchange *symbols*

People use **context** to derive the correct meaning

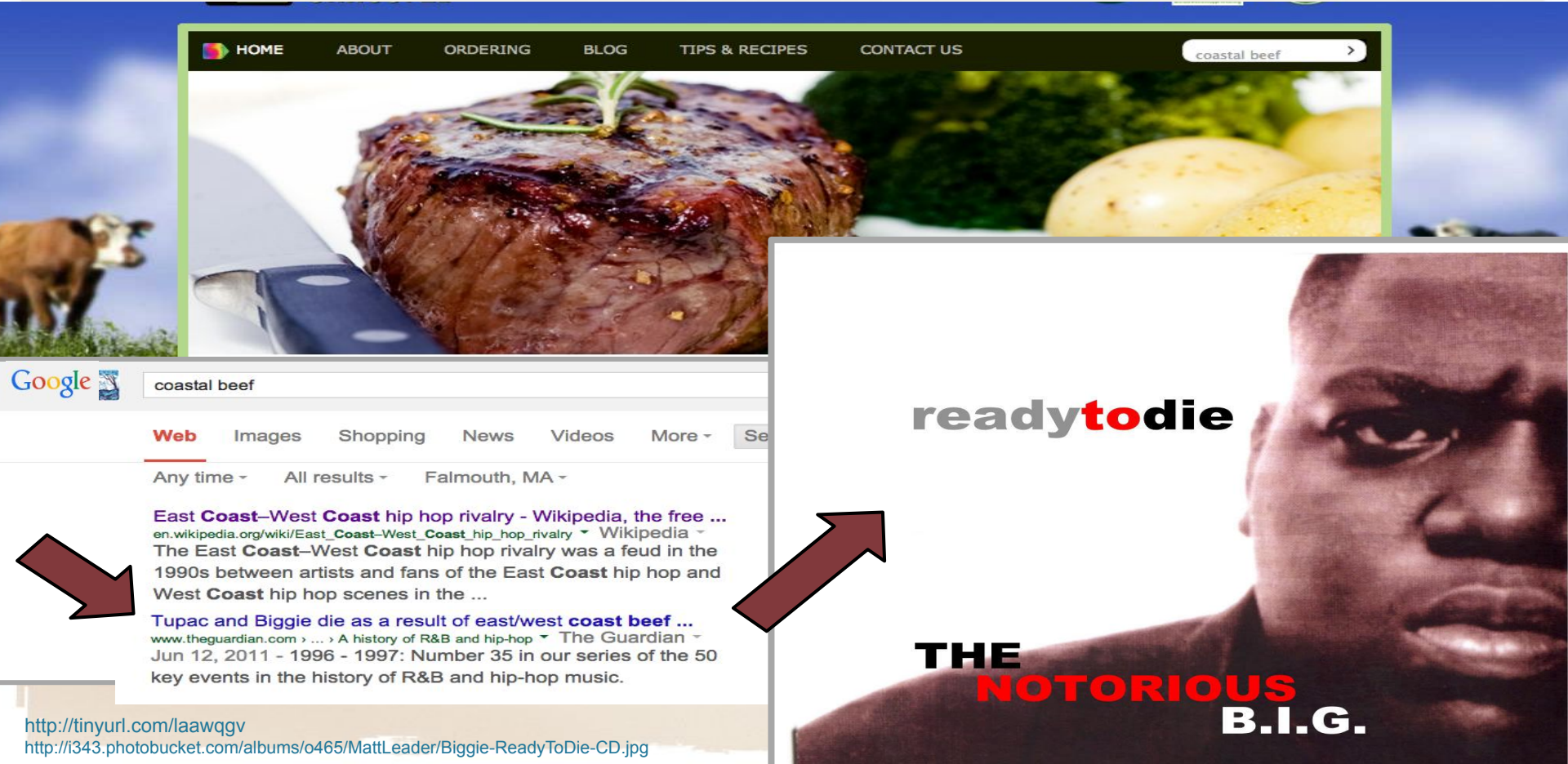
Ogden, C. K., & Richards, I. A. (1923) The Meaning of Meaning

SEARCH: "COASTAL BEEF"



http://www.leftcoastgrassfed.com/

IS **“COASTAL BEEF”** DEADLY?



http://tinyurl.com/laawqgv
http://i343.photobucket.com/albums/o465/MattLeader/Biggie-ReadyToDie-CD.jpg

```
<name>Adam Shepherd</name>
<address>1428 Elm St.</address>
<city>Austin</city>
<state>TX</state>
```

```
<名称>亚当谢泼德</名称>
<地址>1428 榆樹街</地址>
<市>奥斯汀</市>
<國>德州</國>
```

Without a “data dictionary” it’s difficult to know that meaning of the data elements is.

Tags appear in patterns, but meaning is a mystery to a computer

Machines can detect if two URIs are the same.

<name>

<名称>

https://schema.org/name

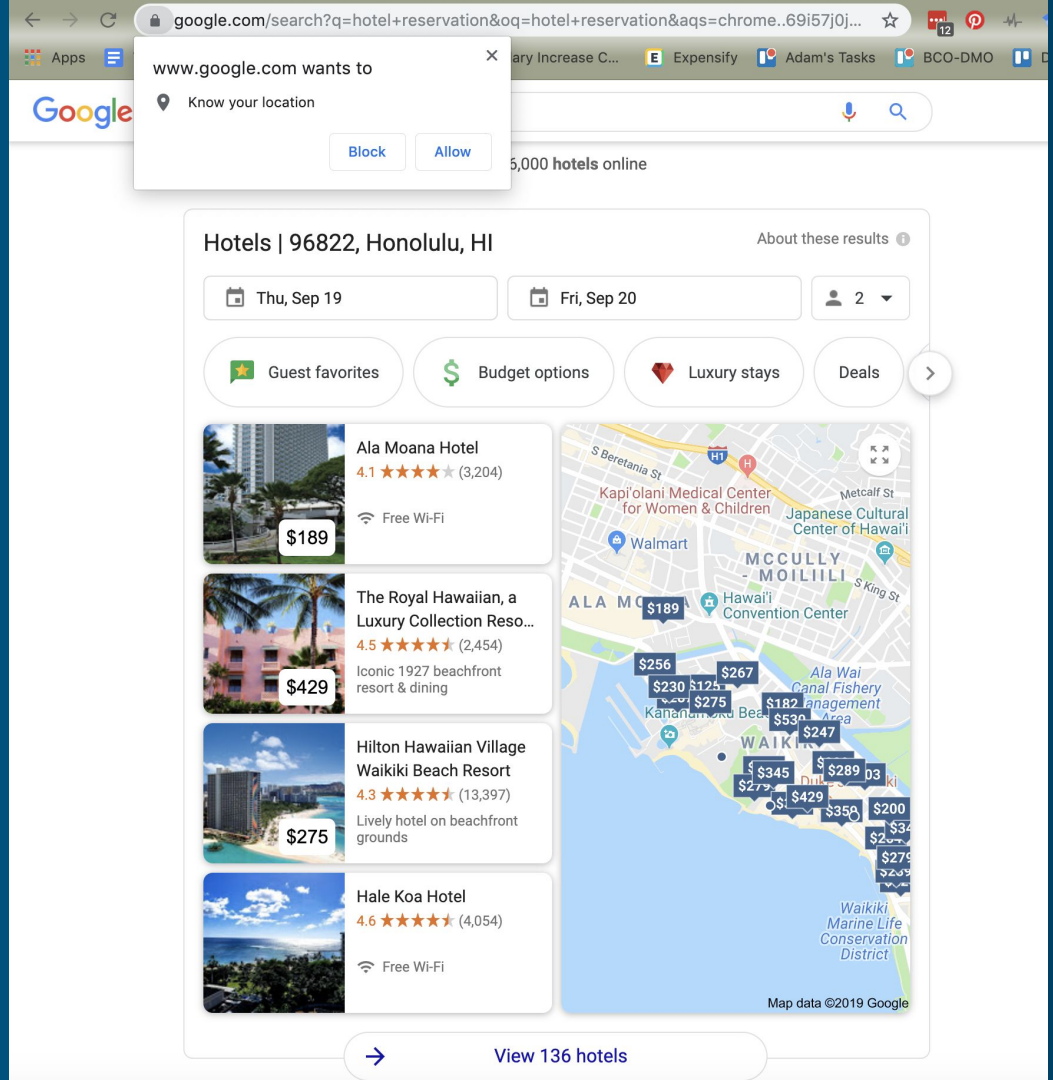
Machines can detect if two URIs are the same.

Adam Shepherd

亚当谢泼德

https://orcid.org/0000-0003-4486-9448

Works now, today



Data on the Web Best Practices

W3C Recommendation 31 January 2017

DISCOVERABILITY

- Provide metadata
- Provide descriptive metadata
- Use persistent URIs as identifiers of datasets
- Use persistent URIs as identifiers within datasets
- Assign URIs to dataset versions and series
- Use Web Standards as the foundation of APIs
- Cite the Original Publication

COMPREHENSION

- Provide metadata
- Provide descriptive metadata
- Provide structural metadata
- Provide data provenance information
- Use locale-neutral data representations
- Reuse vocabularies, preferably standardized ones
- Choose the right formalization level
- Gather feedback from data consumers
- Enrich data by generating new data
- Provide Complementary Presentations

ACCESS

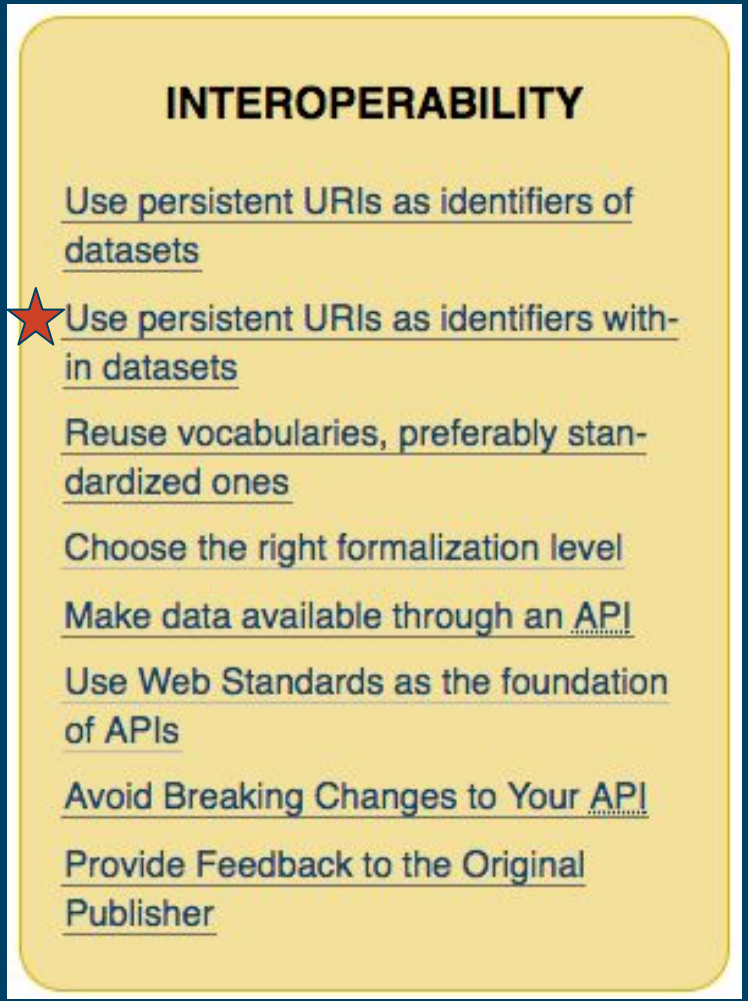
- Provide bulk download
- Provide Subsets for Large Datasets
- Use content negotiation for serving data available in multiple formats
- Provide real-time access
- Provide data up to date
- Make data available through an API
- Use Web Standards as the foundation of APIs
- Provide Complementary Presentations

INTEROPERABILITY

- Use persistent URIs as identifiers of datasets
- Use persistent URIs as identifiers within datasets
- Reuse vocabularies, preferably standardized ones
- Choose the right formalization level
- Make data available through an API
- Use Web Standards as the foundation of APIs
- Avoid Breaking Changes to Your API
- Provide Feedback to the Original Publisher

https://www.w3.org/TR/dwbp/#challenges

W3C
Data on the Web
Best Practices

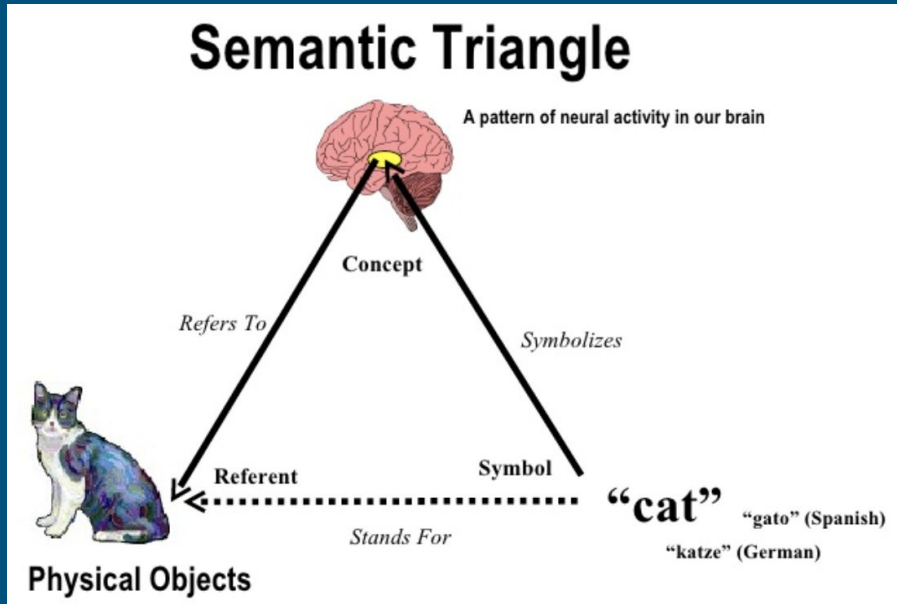


https://www.w3.org/TR/dwbp/#challenges

Making ocean data agreeable (with each other)?

AGREEMENT on identifiers within datasets

- Parameters
- Methods
- Instrumentation
- Etc…



DATABASE

- Programs 44
- Projects 1,085
- Deployments 2,888
- Platforms 596
- Datasets 9,447
- Instruments 487
- Parameters 1,420
- People 2,736
- Affiliations 594
- Funding 93
- Awards 2,044

GEOSPATIAL ACCESS

Parameter: Magnetic_susceptibility

Short Description: Magnetic_susceptibility
Short Name: Magnetic_susceptibility
Official Name: Magnetic susceptibility
Units:
Units External Identifier:
Graphable:
No Data Value:
Maximum Value:
External Identifier:
» skos:exactMatch http://vocab.nerc.ac.uk/collection/P01/current/MAGSSEDM/
Minimum Value:

Description

The degree to which a sediment sample is affected by a magnetic field.

Datasets associated with this parameter

Dataset Name	Brief Description	Project	PI-Supplied Parameter Name
Palau Lake Core Properties	Core Logger Physical Properties for Palau Lakes Sediment Cores	PaPaPro	MSCL_MS_Loop

DATABASE

- Programs 44
- Projects 1,085
- Deployments 2,888
- Platforms 596
- Datasets 9,447
- Instruments 487
- Parameters 1,420
- People 2,736
- Affiliations 594
- Funding 93
- Awards 2,044

GEOSPATIAL ACCESS

Parameter: Magnetic_susceptibility

Short Description: Magnetic_susceptibility
Short Name: Magnetic_susceptibility
Official Name: Magnetic susceptibility
Units:
Units External Identifier:
Graphable:
No Data Value:
Maximum Value:
External Identifier:
» skos:exactMatch http://vocab.nerc.ac.uk/collection/P01/current/MAGSSEDM/
Minimum Value:

1.

Description

The degree to which a sediment sample is affected by a magnetic field.

Datasets associated with this parameter

Dataset Name	Brief Description	Project	PI-Supplied Parameter Name
Palau Lake Core Properties	Core Logger Physical Properties for Palau Lakes Sediment Cores	PaPaPro	MSCL_MS_Loop

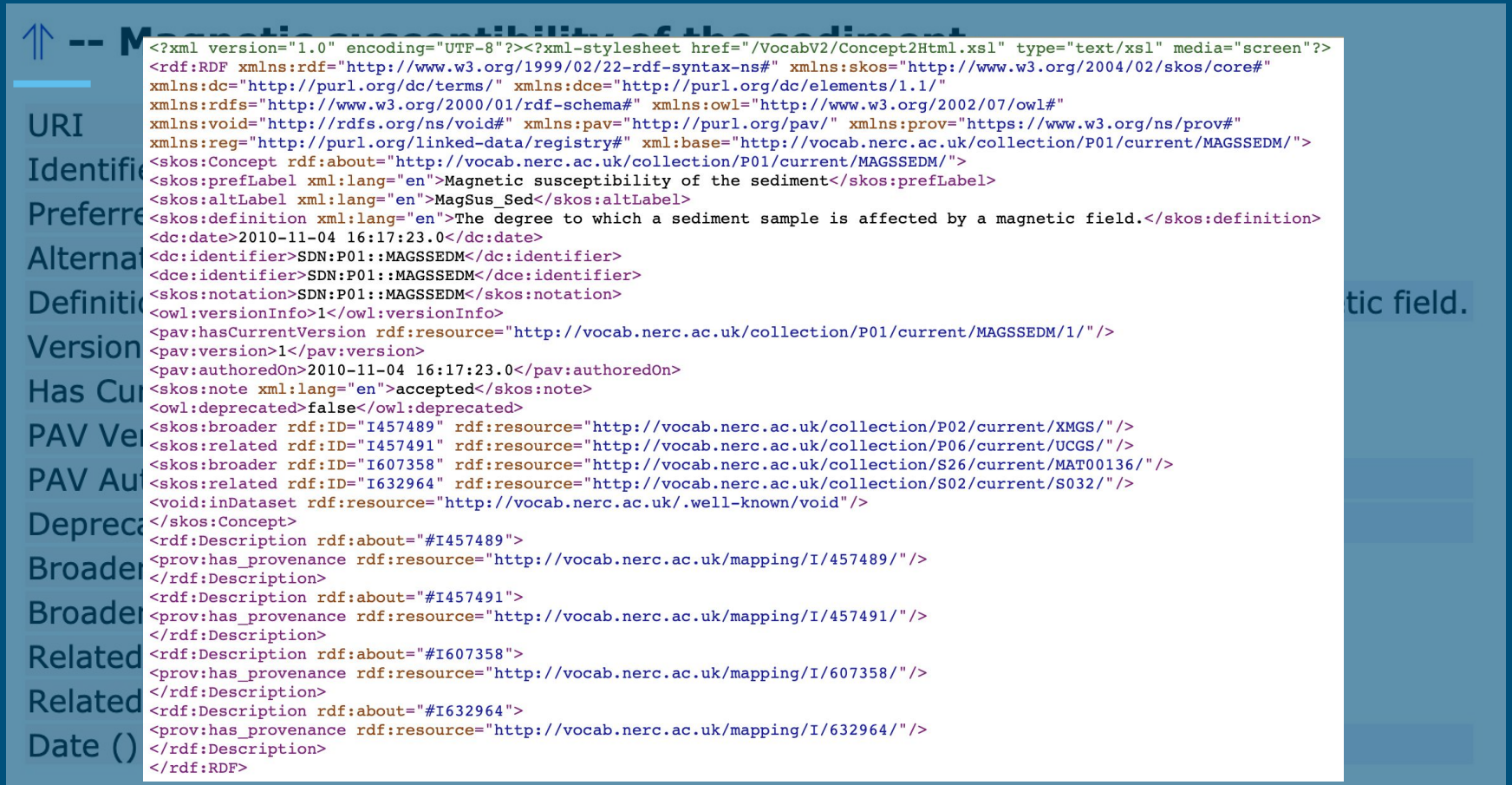
2.

http://vocab.nerc.ac.uk/collection/P01/current/MAGSSEDM/

-- Magnetic susceptibility of the sediment --

URI	http://vocab.nerc.ac.uk/collection/P01/current/MAGSSEDM/
Identifier ()	SDN:P01::MAGSSEDM
Preferred label (en)	**Magnetic susceptibility of the sediment**
Alternative label (en)	MagSus_Sed
Definition (en)	The degree to which a sediment sample is affected by a magnetic field.
Version Info ()	1
Has Current Version	http://vocab.nerc.ac.uk/collection/P01/current/MAGSSEDM/1/
PAV Version ()	1
PAV Authored On ()	2010-11-04 16:17:23.0
Deprecated()	false
Broader	http://vocab.nerc.ac.uk/collection/P02/current/XMGS/
Broader	http://vocab.nerc.ac.uk/collection/S26/current/MAT00136/
Related	http://vocab.nerc.ac.uk/collection/P06/current/UCGS/
Related	http://vocab.nerc.ac.uk/collection/S02/current/S032/
Date ()	2010-11-04 16:17:23.0

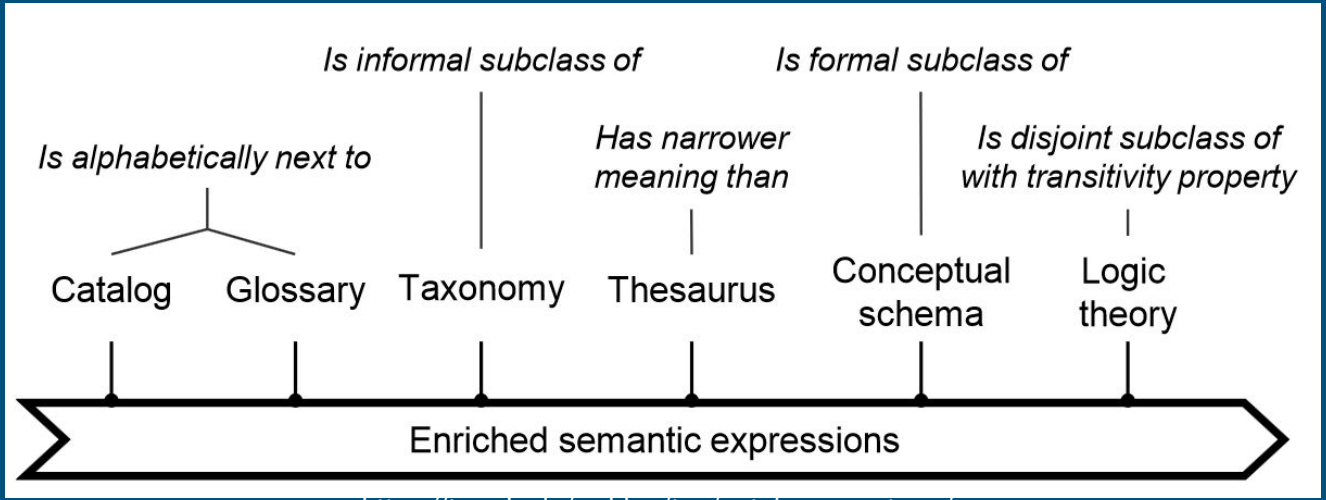
http://vocab.nerc.ac.uk/collection/P01/current/MAGSSEDM/



```
<?xml version="1.0" encoding="UTF-8"?><?xml-stylesheet href="/VocabV2/Concept2Html.xsl" type="text/xsl" media="screen"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:skos="http://www.w3.org/2004/02/skos/core#"
xmlns:dc="http://purl.org/dc/terms/" xmlns:dce="http://purl.org/dc/elements/1.1/"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:void="http://rdfs.org/ns/void#" xmlns:pav="http://purl.org/pav/" xmlns:prov="https://www.w3.org/ns/prov#"
xmlns:reg="http://purl.org/linked-data/registry#" xml:base="http://vocab.nerc.ac.uk/collection/P01/current/MAGSSEDM/">
<skos:Concept rdf:about="http://vocab.nerc.ac.uk/collection/P01/current/MAGSSEDM/">
<skos:prefLabel xml:lang="en">Magnetic susceptibility of the sediment</skos:prefLabel>
<skos:altLabel xml:lang="en">MagSus_Sed</skos:altLabel>
<skos:definition xml:lang="en">The degree to which a sediment sample is affected by a magnetic field.</skos:definition>
<dc:date>2010-11-04 16:17:23.0</dc:date>
<dc:identifier>SDN:P01::MAGSSEDM</dc:identifier>
<dce:identifier>SDN:P01::MAGSSEDM</dce:identifier>
<skos:notation>SDN:P01::MAGSSEDM</skos:notation>
<owl:versionInfo>1</owl:versionInfo>
<pav:hasCurrentVersion rdf:resource="http://vocab.nerc.ac.uk/collection/P01/current/MAGSSEDM/1/"/>
<pav:version>1</pav:version>
<pav:authoredOn>2010-11-04 16:17:23.0</pav:authoredOn>
<skos:note xml:lang="en">accepted</skos:note>
<owl:deprecated>false</owl:deprecated>
<skos:broader rdf:ID="I457489" rdf:resource="http://vocab.nerc.ac.uk/collection/P02/current/XMGS/"/>
<skos:related rdf:ID="I457491" rdf:resource="http://vocab.nerc.ac.uk/collection/P06/current/UCGS/"/>
<skos:broader rdf:ID="I607358" rdf:resource="http://vocab.nerc.ac.uk/collection/S26/current/MAT00136/"/>
<skos:related rdf:ID="I632964" rdf:resource="http://vocab.nerc.ac.uk/collection/S02/current/S032/"/>
<void:inDataset rdf:resource="http://vocab.nerc.ac.uk/.well-known/void"/>
</skos:Concept>
<rdf:Description rdf:about="#I457489">
<prov:has_provenance rdf:resource="http://vocab.nerc.ac.uk/mapping/I/457489/"/>
</rdf:Description>
<rdf:Description rdf:about="#I457491">
<prov:has_provenance rdf:resource="http://vocab.nerc.ac.uk/mapping/I/457491/"/>
</rdf:Description>
<rdf:Description rdf:about="#I607358">
<prov:has_provenance rdf:resource="http://vocab.nerc.ac.uk/mapping/I/607358/"/>
</rdf:Description>
<rdf:Description rdf:about="#I632964">
<prov:has_provenance rdf:resource="http://vocab.nerc.ac.uk/mapping/I/632964/"/>
</rdf:Description>
</rdf:RDF>
```

Controlled Vocabularies

- A list of terms and definitions identified by URIs
- Some governance model for ensuring consistency
- Defined in relation to each other



https://tw.rpi.edu/weblog/tag/ontology-spectrum/

Making ocean data agreeable (with each other)?

Recommendation	Solution	Where is the Solution?
Identify datasets with a persistent URI.	DOIs	**Data Repository***
Identify dataset *contents* with a persistent URI.	Controlled Vocabulary URIs	**Domain Repository**
Reuse vocabularies, preferably standardized ones.	Community	**You!**

* https://www.the-scientist.com/news-opinion/the-push-to-replace-journal-supplements-with-repositories--66296

In Review

1. We need AGREEMENT about dataset contents.
2. Agreements materialized as symbols (URIs) for computers.

Questions?

EXTRA SLIDES

Choosing a Controlled Vocabulary

- Availability
- Quality
  - Governance Model
  - Clarity and Precision
  - Completeness
  - Encoding Format (HTML v. Excel/CSV v. RDF/OWL)
- Community Adoption

Graybeal (2001). https://marinemetadata.tamucc.edu/guides/vocabs/cvchooseimplement/cvchoosing

Works now, today...even for Datasets